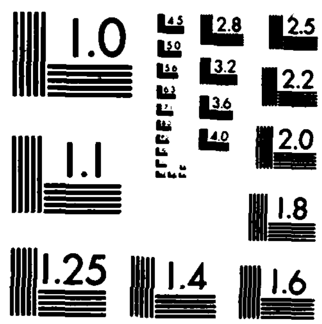
AD-A129 957 TIME SERIES ARMA MODEL IDENTIFICATION BY ESTIMATING INFORMATION(U) TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS E PARZEN MAR 83 TR-N-37 N00014-82-M-2001 1/1

UNCLASSIFIED F/G 12/1 NL

END
DATE
FILMED
7 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

INSTITUTE OF STATISTICS
Phone 713 - 845-3141

TIME SERIES ARMA MODEL IDENTIFICATION BY ESTIMATING INFORMATION

by

Emanuel Parzen
Institute of Statistics
Texas A&M University

Technical Report No. N-37

March 1983

Texas A&M Research Foundation
Project No. 4226T

"Multiple Time Series Modeling and
Time Series Theoretical Statistial Methods"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: N-37	2. GOVT ACCESSION NO.: AD-A129957
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): TIME SERIES ARMA MODEL IDENTIFICATION BY ESTIMATING INFORMATION	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Emanuel Parzen	8. CONTRACT OR GRANT NUMBER(s): ONR N00014-82-MP-2001
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE: March 1983
	13. NUMBER OF PAGES: 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Time series analysis, model identification, information, ARMA scheme, cepstral correlations, exponential spectral model.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Models for time series Y(t) can be formulated as hypotheses concerning the information about Y(t) in other variables. To determine an ARMA model for a univariate time series, this paper synthesizes two approaches: estimation and comparison of various information increments, and non-parametric estimation of the MA(∞) representation to form estimators of the many information numbers that are compared to identify models for a time series.

TIME SERIES ARMA MODEL IDENTIFICATION BY ESTIMATING INFORMATION

Emanuel Parzen

Institute of Statistics
Texas A&M University

Statisticians, economists, and system engineers are becoming aware that to identify models for time series and dynamic systems, information theoretic ideas can play a valuable (and unifying) role. Models for time series $Y(t)$ can be formulated as hypotheses concerning the information about $Y(t)$ given various bases involving past, current, and future values of $Y(\cdot)$ and related time series $X(\cdot)$. To determine sets of variables that are sufficient to forecast $Y(t)$, and especially to determine an ARMA model for $Y(t)$, an approach is presented which estimates and compares various information increments. ~~We discuss~~ how to non-parametrically estimate the MA~~(∞)~~ (infinity) representation, and use it to form estimators of the many information numbers that might compare to identify an ARMA model for a univariate time series.

The author discusses

1. Information Measures

The information approach to model identification formulates a model (or hypothesis about the probability law of random variables or time series) as a hypothesis that an information number is zero. Information measures for random variables are defined in terms of information measures for probability densities. The latter can be regarded as defining "distances" between probability measures.

Let $f(y)$ and $g(y)$ be two probability densities on a real line, $-\infty<y<\infty$. The information divergence of index α of a (model) g from (a true density) f is defined for $\alpha = 1$ (index 1) by

$$I_1(f;g) = \int_{-\infty}^{\infty}\{-\log \frac{g(y)}{f(y)}\}\ f(y)\ dy$$

and for $\alpha>0$ (but $\alpha\neq 1$) by

$$I_\alpha(f;g) = \frac{-1}{1-\alpha}\log \int_{-\infty}^{\infty}\{\frac{g(x)}{f(x)}\}^{1-\alpha} f(x)\ dx \quad .$$

Information divergence of index 1 has a preferred role because it has an important decomposition

$$I_1(f;g) = H(f;g) - H(f)$$

defining

$$H(f;g) = \int_{-\infty}^{\infty}\{-\log g(y)\}f(y)\ dy,$$

$$H(f) = H(f;f) = \int_{-\infty}^{\infty}\{-\log f(y)\}\ f(y)\ dy \quad .$$

We call $H(f;g)$ the cross-entropy of f and g, and $H(f)$ the entropy of f. Information divergence of index 1 is usually referred to just as information divergence $I(f;g)$.

The information $I(Y|X)$ about a continuous random variable Y in a continuous random variable X is defined by

$$I(Y|X) = I(f_{Y|X};\ f_Y) = E_X I\ (f_{Y|X=x};f_Y).$$

The entropy of Y and conditional entropy of Y given X are defined by

$$H(Y) = H(f_Y)$$

$$H(Y|X) = H(f_{Y|X}) = E_X H(f_{Y|X=x}) \quad .$$

One can establish a fundamental decomposition:

$$I(Y|X) = H(Y) - H(Y|X).$$

The most fundamental concept used in identifying models by estimating information is $I(Y|X_1;\ X_1,\ X_2)$, the information about Y in X_2 conditional on X_1; it is defined

$$\text{(I)}\quad I(Y|X_1;X_1,X_2) = H(f_{Y|X_1}) - H(f_{Y|X_1,X_2}) = H(Y|X_1) - H(Y|X_1,X_2) \quad .$$

A fundamental formula to evaluate $I(Y|X_1;X_1,X_2)$ is

$$\text{(II)}\quad I(Y|X_1;X_1,X_2) = I(Y|X_1,X_2) - I(Y|X_1) \quad .$$

When X and Y are jointly normal random variables $f_{Y|X=x}(y)$ is a normal distribution whose variance (which does not depend on x) is denoted $\Sigma(Y|X)$. The variance of Y is denoted $\Sigma(Y)$. The entropy and conditional entropy of Y are

$$H(Y) = \frac{1}{2}\log \Sigma(Y) + \frac{1}{2}(1+\log 2\pi)$$

$$H(Y|X) = \frac{1}{2}\log \Sigma(Y|X) + \frac{1}{2}(1+\log 2\pi) \quad .$$

The information about Y in X when X and Y are bivariate normal, with correlation coefficient ρ, can be expressed

$$\text{(III)}\quad I(Y|X) = -\frac{1}{2}\log \Sigma^{-1}(Y)\Sigma(Y|X) = -\frac{1}{2}\log(1-\rho^2).$$

When Y and X are jointly multivariate normal random vector, let Σ denote a covariance matrix. One can show that

$$\text{(IV)}\quad I(Y|X) = (-\frac{1}{2})\log\det \Sigma^{-1}(Y)\Sigma(Y|X) = (-\frac{1}{2})\ \text{sum log eigenvalues}\ \Sigma^{-1}(Y)\Sigma(Y|X).$$

Research supported by Office of Naval Research under contract no. N00014-82-MP-2001.

To illustrate the information approach to model identification (or determining relations between random variables) consider the general problem of testing the hypothesis H_0: X and Y are independent. One could express H_0 in any one of the following equivalent ways:

H_0: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y;

H_0: $f_{Y|X=x}(y) = f_Y(y)$ for all x and y;

H_0: $I(f_{X,Y};\, f_X f_Y) = 0$;

H_0: $I(Y|X) = 0$.

The information approach to testing H_0 is to form an estimator $\hat{I}(Y|X)$ of $I(Y|X)$, and test whether it is significantly different from zero. One can distinguish several types of estimators of $I(Y|X)$: (a) fully parametric, (b) fully non-parametric; (c) functionally parametric which uses functional statistical inference smoothing techniques to estimate $I(Y|X)$ [see Woodfield (1982)].

An example of fully parametric estimators arises when one assumes X and Y are bivariate normal with correlation coefficient ρ. Given a random sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ a fully parametric estimator of $I(Y|X)$ is the maximum likelihood estimator

$$\hat{I}(Y|X) = -\frac{1}{2}\log(1-\hat{\rho}^2)$$

where $\hat{\rho}$ is the sample correlation coefficient.

2. Information and Memory Approach to Time Series Model Identification

The approach to time series analysis developed by Parzen distinguishes four general types of time series models:

1. No memory or white noise
2. Short memory or stationary
3. Long memory (or non-stationary)

3a. Long memory: transform to short memory

3b. Long memory: long memory plus short memory.

Memory type can be defined in terms of the information numbers I_m

$$I_m = I(Y|Y_{-1},\ldots,Y_{-m})$$
$$= I(Y(t)|Y(t-1),\ldots,Y(t-m)) \quad ;$$

in words, I_m is the information about a time series $Y(t)$ at time t in the m most recent values $Y(t-1),\ldots,Y(t-m)$. Let Y^- denote the infinite past $Y(t-1), Y(t-2),\ldots$. As m tends to ∞, I_m tends to

$$I_\infty = I(Y|Y^-) = I(Y(t)|Y(t-1),\ldots) \quad .$$

We define a time series $Y(t)$, $t=0,\pm1,\ldots$ to be:

no memory if $I_\infty = 0$

short memory if $0 < I_\infty < \infty$

long memory if $I_\infty = \infty$.

The models we build for a time series depend on its memory type. A model corresponds to a transformation of the time series to a no memory (white noise) series. Therefore a no memory (white noise) time series requires no further modeling, although one may be interested in determining such statistical characteristics as the mean, variance, and probability distribution.

A short memory time series $Y(t)$ is modeled by an invertible filter which transforms it to white noise:

$$Y(t) \longrightarrow \boxed{\text{innovations filter } g_\infty} \longrightarrow \epsilon(t) = Y^\nu(t)$$

where $Y^\nu(\cdot)$ is the <u>innovation</u> series, or series of infinite memory one-step ahead prediction errors, defined by [using ν to connote "what's new"]

$$Y^\nu(t) = Y(t) - Y^\mu(t) \quad .$$

The predictor $Y^\mu(t)$ is denoted

$$Y^\mu(t) = E[Y(t)|Y(t-1),\, Y(t-2),\, \ldots]$$
$$= (Y|Y_{-1},\ldots,Y_{-n},\ldots)(t) \quad .$$

We use μ as the superscript for a predictor to indicate that it is an averaging operator.

The infinite memory mean square prediction error is defined as the normalized variance

$$\sigma_\infty^2 = E[|Y^\nu(t)|^2] \div E[|Y(t)|^2] \quad .$$

The appropriateness of normalizing is justified by the formula for information:

$$I_\infty = -\frac{1}{2}\log\sigma_\infty^2$$

if a time series $Y(t)$, $t=0,\pm1,\ldots$ is a zero mean Gaussian stationary time series. Its probability law can be described by the covariance function

$$R(v) = E[Y(t)Y(t+v)] \quad ,$$

and correlation function

$$\rho(v) = \frac{R(v)}{R(0)} = \mathrm{Corr}\,[Y(t),Y(t+v)] \quad .$$

Alternatively the probability law of $Y(\cdot)$ can be described by the spectral density function f which is defined by

$$f(\omega) = \sum_{v=-\infty}^{\infty} e^{-2\pi i v\omega}\rho(v), \qquad 0\le\omega\le1$$

when $\sum_{v=-\infty}^{\infty}|\rho(v)|<\infty$. The frequency variable ω is usually assumed to vary in the interval $-0.5\le\omega\le0.5$. But only the interval $0\le\omega\le0.5$ has physical significance. We prefer the interval $0\le\omega\le1$ for mathematical reasons.

Perhaps the most insightful way to model a short memory time series is by representing it, or approximating it, by an ARMA(p,q) scheme:

$$Y(t) + \alpha_p(1)Y(t-1)+\ldots+\alpha_p(p)Y(t-p)$$
$$= \epsilon(t) + \beta_q(1)\epsilon(t-1)+\ldots+\beta_q(q)\epsilon(t-q)$$

where the polynomials

$$g_p(z) = 1+\alpha_p(1)z+\ldots+\alpha_p(p)\ z^p$$
$$h_q(z) = 1+\beta_q(1)z+\ldots+\beta_q(q)\ z^q$$

are chosen so that all their roots in the complex z-plane are in the region $\{z:|z|>1\}$ outside the unit circle. Then $g_p(z)$ and $h_q(z)$ are the transfer functions of invertible filters. $\varepsilon(t)$ is assumed to be a white noise time series which we identify with the innovations $\varepsilon(t) = Y^\nu(t)$;

$$\sigma^2_{p,q} = E[\varepsilon^2(t)] \div E[Y^2(t)]$$

is an estimator of σ^2_∞. The spectral density of an ARMA (p,q) scheme is

$$f_{p,q}(\omega) = \sigma^2_{p,q} \frac{|h_q(e^{2\pi i\omega})|^2}{|g_p(e^{2\pi i\omega})|^2}$$

The process of identifying ARMA(p,q) schemes which are adequate (and parsimonious) approximating models for a time series can be studied by determining information characterizations of when the exact (or true) model is an AR(p) or ARMA(p,q).

Let $\Sigma(Y|Y_{-1},\ldots,Y_{-p}, Y^\nu_{-1},\ldots,Y^\nu_{-q})$ denote the mean square prediction error of Y(t) when predicted by $Y(t-1),\ldots,Y(t-p)$; $Y^\nu(t-1),\ldots,Y^\nu(t-q)$, or equivalently the conditional variance of Y(t) given $Y(t-1),\ldots,Y(t-p)$, $Y^\nu(t-1),\ldots, Y^\nu(t-q)$. Normalize it to form

$$\sigma^2_{p,q} = \Sigma^{-1}(Y)\Sigma(Y|Y_{-1},\ldots,Y_{-p}, Y^\nu_{-1},\ldots,Y^\nu_{-q})$$
$$I_{p,q} = -\frac{1}{2}\log \sigma^2_{p,q} \quad .$$

The information difference between $Y_{-1},\ldots,Y_{-p}$, $Y^\nu_{-1},\ldots,Y^\nu_{-q}$ and Y^- for prediction of Y(t) satisfies

$$I(Y|Y_{-1},\ldots,Y_{-p},Y^\nu_{-1},\ldots,Y^\nu_{-q};Y^-) = I_\infty - I_{p,q} \quad .$$

The following two hypotheses are equivalent:

H_0: $Y(\cdot)$ is ARMA(p,q)

H_0: $I_\infty - I_{p,q} = 0$.

3. Information Calculation for ARMA Schemes

Given a sample $\{Y(t),\ t=1,2,\ldots T\}$, we would like to estimate, for many values of p,q, the information differences (assuming normality)

$$I_\infty - I_{p,q} = -\frac{1}{2}\log \sigma^2_\infty - \{-\frac{1}{2}\log \sigma^2_{p,q}\} \ .$$

We need to estimate σ^2_∞ and $\sigma^2_{p,q}$. To understand the method we would like to propose, let us first discuss how to compute the true value of $\sigma^2_{p,q}$. The MA(∞), or infinite order moving average, representation of Y(t) will play a central role:

$$Y(t) = Y^\nu(t) + \beta_1 Y^\nu(t-1) + \beta_2 Y^\nu(t-2) + \ldots \ .$$

Note that $E[|Y(t)|^2] = E[|Y^\nu(t)|^2]\ \{1+\beta_1^2+\ldots\}$ so that

$$1 = \sigma^2_\infty\ \{1 + \beta_1^2+\beta_2^2 + \ldots\} \quad .$$

The correlations $\rho(v)$ can be computed by

$$\rho(v) = \sigma^2_\infty\ \{\beta_v + \beta_1\ \beta_{v+1} + \ldots\} \quad .$$

By using matrix sweep operations on the joint covariance matrix of $Y,Y_{-1},\ldots,Y_{-p}, Y^\nu_{-1},\ldots,Y^\nu_{-q}$ one can determine (in a stepwise manner) the conditional variance $\Sigma(Y|Y_{-1},\ldots,Y_{-p}, Y^\nu_{-1},\ldots, Y^\nu_{-q})$ required to compute the information $I_{p,q}$.

We illustrate the approach being proposed in the case p=1, q=1. The covariance matrix of Y, Y_{-1}, Y^ν_{-1} is

$$\Sigma = \begin{bmatrix} 1 & \rho(1) & \sigma^2_\infty\beta_1 \\ \rho(1) & 1 & \sigma^2_\infty \\ \sigma^2_\infty\beta_1 & \sigma^2_\infty & \sigma^2_\infty \end{bmatrix}$$

Sweep Σ on Y_{-1} to obtain

$$\Sigma_1 = \begin{bmatrix} 1-\rho^2(1) & \rho(1) & \sigma^2_\infty\ \beta_1-\rho(1)) \\ -\rho(1) & 1 & -\sigma^2_\infty \\ \sigma^2_\infty(\beta_1-\rho(1)) & \sigma^2_\infty & \sigma^2_\infty(1-\sigma^2_\infty) \end{bmatrix}$$

Sweep Σ on Y^ν_{-1} to obtain

$$\Sigma_2 = \begin{bmatrix} 1-\sigma^2_\infty\beta_1^2 & \rho(1)-\sigma^2_\infty\beta_1 & \beta_1 \\ \rho(1)-\sigma^2_\infty\beta_1 & 1-\sigma^2_\infty & 1 \\ -\beta_1 & -1 & (\sigma^2_\infty)^{-1} \end{bmatrix}$$

Sweep Σ_1 on Y^ν_{-1} or sweep Σ_2 on Y_{-1} to obtain a matrix which we write in the following form:

$$\begin{bmatrix} (1-\rho^2(1)) - \frac{(\beta_1-\rho(1))^2\sigma^2_\infty}{1-\sigma^2_\infty} & \frac{\rho(1)-\sigma^2_\infty\beta_1}{1-\sigma^2_\infty} & \frac{\beta_1-\rho(1)}{1-\sigma^2_\infty} \\ -\frac{\rho(1)-\sigma^2_\infty\beta_1}{1-\sigma^2_\infty} & \frac{1}{1-\sigma^2_\infty} & \frac{-1}{1-\sigma^2_\infty} \\ -\frac{\beta_1-\rho(1)}{1-\sigma^2_\infty} & \frac{-1}{1-\sigma^2_\infty} & \frac{1}{\sigma^2_\infty(1-\sigma^2_\infty)} \end{bmatrix}$$

We conclude that

$$\Sigma(Y|Y_{-1}) = 1-\rho^2(1),\ (Y|Y_{-1})(t) = \rho(1)\ Y_{-1}(t)$$
$$\Sigma(Y|Y^\nu_{-1}) = 1 - \sigma^2_\infty\beta_1^2 \quad (Y|Y^\nu_{-1})(t) = \beta_1 Y^\nu_{-1}(t)$$
$$\Sigma(Y|Y_{-1},Y^\nu_{-1}) = (1-\rho^2(1)) - \frac{(\beta_1-\rho(1))^2\sigma^2_\infty}{1-\sigma^2_\infty}$$

$$(Y|Y_{-1},Y^{\nu}_{-1})(t) = \frac{\rho(1)-\sigma_\infty^2\beta_1}{1-\sigma_\infty^2} Y_{-1}(t) + \frac{\beta_1-\rho(1)}{1-\sigma_\infty^2} Y^{\nu}_{-1}(t) \quad .$$

These coefficients of $Y_{-1}(t)$ and $Y^{\nu}_{-1}(t)$ can be used as initial (or perhaps even final) values for an efficient parameter estimation algorithm for an ARMA(1,1).

As a check on these formulas, note that for an MA(1), $Y(t) = \varepsilon(t)+b\,\varepsilon(t-1)$, $\beta_1 = b$, $\sigma_\infty^2 = (1+b^2)^{-1}$, $\rho(1) = b/(1+b^2)$. The coefficients of $Y_{-1}(t)$ and $Y^{\nu}_{-1}(t)$ in the predictor are respectively 0 and b.

For a numerical illustration of these formulas, consider the ARMA(1,1) model $Y(t)-aY(t-1) = Y^{\nu}(t)+bY^{\nu}(t-1)$. Then $\beta_1 = a+b$. $1=\sigma_\infty^2(1+(\beta_1^2/(1-a^2)))$, $\rho(1) = [\beta_1+(\beta_1^2 a/(1-a^2))]\sigma_\infty^2$. For $a=b=0.5$, $\beta_1 = 1$, $\sigma_\infty^2 = 3/7$, $\rho(1) = 5/7$. The general formulas yield the values assumed in the model.

To test whether a time series $Y(\cdot)$ obeys an ARMA(1,1), form

$$I(Y|Y_{-1},Y^{\nu}_{-1};Y^-)=\frac{1}{2}\log\left\{\frac{1-\rho^2(1)}{\sigma_\infty^2} - \frac{\{\beta_1-\rho(1)\}^2}{1-\sigma_\infty^2}\right\}$$

This information number equals 0 if the time series obeys any one of the schemes AR(1), MA(1), or ARMA(1,1). The information numbers for an AR(1) and MA(1) are respectively

$$I(Y|Y_{-1};Y^-) = \frac{1}{2}\log\left\{\frac{1-\rho^2(1)}{\sigma_\infty^2}\right\} \quad ;$$

$$I(Y|Y^{\nu}_{-1};Y^-) = \frac{1}{2}\log\left\{\frac{1}{\sigma_\infty^2} - \beta_1^2\right\} \quad .$$

One accepts H_0: $Y(\cdot)$ is ARMA(1,1) if the last two information numbers are different from zero, but $I(Y|Y_{-1}, Y^{\nu}_{-1}; Y^-) = 0$.

For the ARMA(1,1) model $Y(t) - 0.5\,Y(t-1) = Y^{\nu}(t) + 0.5\,Y^{\nu}(t-1)$,

$$I(Y|Y_{-1};\,Y^-) = \frac{1}{2}\log\frac{8}{7} = .067$$

$$I(Y|Y^{\nu}_{-1};\,Y^-) = \frac{1}{2}\log\frac{4}{3} = .143 \quad .$$

When information $I_{p,q}$ is estimated from a sample of size T, a penalty term $(1+p+q)/T$ is subtracted from the estimated information $I_{p,q}$ in the Akaike information approach. If .067 were an estimated value of $I(Y|Y_{-1};\,Y^-)$ it would be regarded as significantly different from zero if $.067 - (2/T) \geq 0$, which is true for $T \geq 30$.

To identify the best orders p,q of approximating ARMA(p,q) one could use subset regression techniques to steer the calculation of $I_{p,q}$. Alternatively one could compute the information numbers of AR(p), MA(q), ARMA(p,q) for p,q = 1, ...,M (a specified upper limit). Subtract from estimated information number a penalty $(1+p+q)/T$. Then sort the array of penalized estimated information numbers $I_{p,q}$ to determine the orders (p,q) of schemes with the largest amount of information (and which therefore minimize $I_\infty - I_{p,q}$ and correspond to best approximating ARMA schemes by this measure of divergence between probability distributions).

4. Nonparametric Estimation of MA(∞) Representation

An information approach to computing $I_{p,q}$ and thus identifying best fitting schemes has been described which is based on estimating the coefficients of the MA(∞) representation. Two possible methods for non-parametric MA(∞) estimation are described in this section: (1) approximating long autoregressive schemes; (2) cepstral correlations. The two methods may be used simultaneously for greater confidence in the results obtained. Both methods require further theoretical investigation [compare Bhansali (1982)].

Denote the MA(∞) representation of Y(t) by

$$Y(t) = b(0)\,Y^{\nu}(t) + b(1)\,Y^{\nu}(t-1) + \ldots$$

where b(0) = 1. Denote the AR(∞) representation by

$$a(0)\,Y(t) + a(1)\,Y(t-1) + \ldots = Y^{\nu}(t)$$

where a(0) = 1.

The approximating long autoregressive scheme estimates the AR(∞) representation of a time series $Y(\cdot)$ by a finite order AR(p) scheme

$$Y(t) + a_p(1)Y(t-1)+\ldots+a_p(p)Y(t-p) = \varepsilon(t)$$

whose order p is determined by an order determining scheme [such as AIC, due to Akaike, or CAT, due to Parzen]. The generating functions

$$h_\infty(z) = 1 + b(1)z + b(2)z^2+\ldots$$

$$g_\infty(z) = 1\ a(1)z + a(2)z^2+\ldots$$

$$g_p(z) = 1 + a_p(1)\,z+\ldots+a_p(p)\,z^p$$

satisfy

$$g_\infty(z)\,h_\infty(z) = 1.$$

One can solve recursively for b(j) using the recursion

$$a(0)\,b(k) + a(1)\,b(k-1)+\ldots+a(k)\,b(0) = 0.$$

When $g_\infty(z)$ is approximated by $g_p(z)$, one replaces a(k) by $a_p(k)$; note that $a_p(k) = 0$ for k>p. The approximating autoregressive method of estimating the MA(∞) representation often yields reasonable results in practice. However it is difficult to study its properties theoretically.

The cepstral correlation method is available for

short memory time series; then $\log f(\omega)$ is integrable, and can be used to compute I_∞ using the fundamental formula (due to Kolmogorov and Szego)

$$\log \sigma_\infty^2 = \int_0^1 \log f(\omega)\, d\omega \quad .$$

The cepstral correlations are defined by, for $v=0, \pm 1, \ldots,$

$$\psi(v) = \int_0^1 e^{2\pi i \omega v} \log f(\omega)\, d\omega \quad .$$

The name "cepstral correlations" is intended to connote that $\psi(v)$ is the Fourier transform of $\log f(\omega)$. However the sequence $\{\psi(v)\}$ does not share an essential property of the sequence $\{\rho(v)\}$ of correlations; the cepstral-correlations are not non-negative definite since $\log f(\omega)$ is not non-negative. Define

$$\Psi(z) = \sum_{k=1}^{\infty} \psi(k)\, z^k \; , \; \Psi^*(z) = \sum_{k=1}^{\infty} \psi(-k)\, z^{-k} \; .$$

Then

$$f(\omega) = \sigma_\infty^2 \, |h\,(e^{2\pi i \omega})|^2$$

and

$$\log f(\omega) = \Psi_0 + \Psi(e^{2\pi i \omega}) + \Psi^*(e^{2\pi i \omega}) \quad .$$

A very important relation [which goes back to the dawn of modern time series analysis, due to Kolmogorov (1939)] is

$$h_\infty(z) = \exp \Psi(z).$$

One can obtain an explicit formula for $b(k)$ in terms of $\Psi(k)$; thus Janacek (1982) writes

$$b(1) = \psi(1),$$

$$b(2) = \psi(2) + \psi^2(1)/2!,$$

$$b(3) = \psi(3) + \psi(1)\, \psi(2) + \psi^3(1)/3! \quad .$$

A more useful representation of the formula for $b(k)$ in terms of $\psi(k)$ has been given by Pourahmadi (1982):

$$b(n+1) = \sum_{j=0}^{n} \left(1 - \frac{j}{n+1}\right) \psi(n+1-j)\, b(j).$$

We outline Pourahmadi's proof; differentiate with respect to z the relation $h_\infty = \exp \Psi$. Obtain $h_\infty' = h_\infty \Psi'$; explicitly

$$\sum_{n=1}^{\infty} nb(n)\, z^{n-1} = \left\{ \sum_{n=0}^{\infty} b(n)\, z^n \right\} \left\{ \sum_{n=1}^{\infty} n\psi(n)\, z^{n-1} \right\}$$

or

$$\sum_{n=0}^{\infty} (n+1) b(n+1) z^n = \left\{ \sum_{n=0}^{\infty} b(n) z^n \right\}$$

$$\left\{ \sum_{n=0}^{\infty} (n+1)\psi(n+1)\, z^n \right\} \quad .$$

Therefore

$$(n+1)\, b(n+1) = \sum_{k=0}^{n} (k+1)\psi(k+1)\, b(n-k)$$

$$= \sum_{j=0}^{n} b(j)\, (n+1-j)\, \psi(n+1-j) \quad .$$

Divide by $n+1$ to obtain the desired conclusion.

Pourahmadi (1982) also states a recursive formula for computation of the AR(∞) coefficients $a(k)$ from $\psi(k)$:

$$a(n+1) = -\sum_{j=0}^{n} \left(1 - \frac{j}{n+1}\right) \psi(n+1-j) a(j) \quad .$$

The properties of cepstral correlations can be understood by examining their values in the case of an AR(1); then

$$f(\omega) = \sigma_\infty^2 \left|1 - \rho e^{2\pi i \omega}\right|^{-2}$$

where $|\rho| < 1$. Then, for $k \geq 1$,

$$\psi(k) = \int_0^1 - \log\, \{1 - \rho\, e^{-2\pi i \omega}\} e^{2\pi i k \omega}$$

$$= \frac{1}{k}\, \rho^k \quad .$$

The rate of decay of $k\psi(k)$, $k=1,2,\ldots,$ is a measure of the memory of the time series.

To estimate $\psi(k)$ from a sample $Y(t)$, $t=1,\ldots,T$ one could take the logarithm of the sample spectral density (computed for $\omega = k/Q$, where one should choose $Q \geq 2T$)

$$\tilde{f}(\omega) = \left| \sum_{t=1}^{T} Y(t) \exp\,(2\pi i \omega t) \right|^2 \div \sum_{t=1}^{T} |Y(t)|^2$$

or a smoothed estimator $\hat{f}(\omega)$ of $f(\omega)$. Then

$$\hat{\psi}(v) = \frac{1}{Q} \sum_{k=0}^{Q-1} \log \tilde{f}\left(\frac{k}{Q}\right) \exp\,(2\pi i v k/Q)$$

A convenient formula for $\hat{f}(\omega)$ is the windowed periodogram of bandwidth $1/T$ defined by

$$\hat{f}(\omega) = \sum_{|v|<T} k\left(\frac{v}{T}\right) \hat{\rho}(v) \exp\,(2\pi i v \omega)$$

where $\hat{\rho}(v)$ is the sample correlation function computed by

$$\hat{\rho}(v) = \frac{1}{Q} \sum_{k=0}^{Q-1} \tilde{f}\left(\frac{k}{Q}\right) \exp\,(2\pi i v k/Q)$$

and $k(t)$ is a suitable kernel (providing non-negative estimators) such as the Parzen window

$$\begin{aligned} k(t) &= 1 - 6t^2 + 6t^3 \; , & |t| &\leq 0.5 \; , \\ &= 2\,(1 - |t|)^3 \; , & 0.5 \leq |t| &\leq 1 \; , \\ &= 0 \; , & 1 &\leq |t| \; . \end{aligned}$$

A kernel with superior properties (but not necessarily non-negative estimates) is the spline-equivalent window [Parzen (1958), Cogburn and Davis (1974), Wahba (1980)]

$$k(t) = \frac{1}{1 + t^{2r}}$$

where r is usually chosen to equal 2 or 4.

An obvious moral of the foregoing formulas is that modern time series model identification requires the scientist to integrate time domain and frequency domain techniques. The cepstral correlations approach to ARMA model

identification also may provide a unification of ARMA models and the exponential spectral models introduced by Bloomfield (1973).

5. Conclusion

Given a sample of time series, one should estimate its correlations $\rho(v)$ and cepstral correlation $\psi(v)$ through Fast Fourier transformation from the sample spectral density $\tilde{f}(\omega)$ and its logarithm log $\tilde{f}(\omega)$.

Using the estimated correlations, the Yule-Walker equations are solved to estimate innovation variances σ_m^2, $m=1,2,\ldots$. Order determining criteria, such as AIC and CAT, are applied to this sequence to determine orders $\hat{m}$ of approximating AR schemes, to determine the memory type of the time series [Parzen (1982)], and to form autoregressive estimators of $f(\omega)$, log $f(\omega)$, and $\psi(v)$.

When a time series is classified as short memory the estimated cepstral correlations are used to form the MA(∞) coefficients b(k). They are used to form information numbers (via sweep or subset regression procedures) for determining best fitting ARMA schemes, and the corresponding ARMA spectral density estimator.

We do not believe that spectral estimation is a non-parametric procedure to be conducted independently of model identification. The final form of spectral estimator should be based on an identification of the type (AR, MA, or ARMA) of the whitening filter of a short memory time series.

Statistical computing has a vital role in time series analysis in two important ways: (1) to rapidly make available to the broader scientific community new algorithms for time series analysis; (2) to make old theoretical ideas of time series analysis practically useful and to stimulate the integration of old and new techniques of time series analysis.

For other aspects of the role of entropy and information measures in model identification, see Akaike (1977) and IFAC (1982). For modeling of multiple time series, see Parzen and Newton (1980), Newton (1983), and Cooper and Wood (1982). A review (and power study) of some standard statistical procedures for determining the orders p and q of an ARMA scheme is given by Clarke and Godolphin (1982).

REFERENCES

Akaike, H. (1977). On entropy maximization principle, Application of Statistics, P. R. Krishnaiah, ed., North-Holland: Amsterdam, 27-41.

Bhansali, R. J. (1982). A comparison of the Wiener-Kolmogorov and the Yule-Walker methods of fitting autoregressions for prediction. Time Series Analysis: Theory and Practice I, ed. O. D. Anderson, North Holland: Amsterdam, 31-52.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. Biometrika 60, 217-226.

Clarke, B. R. and Godolphin, E. J. (1982). Comparative Power Studies for Goodness of Fit Tests of Time Series Models, Journal of Time Series Analysis, 3, 141-151.

Cogburn, R. and Davis, H. T. (1974). Periodic splines and spectral estimation. Annals of Statistics, 2, 1108-1126.

Cooper, D. M. and Wood, E. F. (1982). Identifying multivariate time series models. Journal of Time Series Analysis, 3, 153-164.

IFAC (1982). Symposium on Identification and System Parameter Identification, Arlington, Virginia, June 7-11, 1982. Session on "New Ideas in System Identification" emphasizing information theory and entropy function approaches.

Janacek, G. (1982). Determining the degree of differencing for time series via the log spectrum. Journal of Time Series Analysis, 3, 177-183.

Kolmogorov, A. (1939). Interpolation and extrapolation of stationary random sequences (in French). Comptes Rendus Ac. Sci. (Paris), 208, 2043-2045.

Newton, H. Joseph (1983). An Introduction to the Methods of Time Series Analysis in the Time and Frequency Domains: A User's Guide to the TIMESBOARD Computing Library. Manuscript.

Parzen, E. (1958). "On asymptotically efficient consistent estimates of the spectral density function of a stationary time series" J. Roy. Statist. Soc., B., 20, 303-322.

Parzen, E. (1982). Time Series Model Identification and Prediction Variance Horizon, Applied Time Series Analysis II, ed. D. Findley, Academic Press: New York, 415-447.

Parzen, E. and Newton, H. J. (1980). Multiple Time Series Modeling, II Multivariate Analysis - V, edited by P. Krishnaiah, North Holland: Amsterdam, 181-197.

Pourahmadi, Mohsen (1982). Exact factorization of the spectral density and its application to forecasting and time series analysis. Manuscript (in press).

Wahba, Grace (1980). Automatic smoothing of the log periodogram, Journal of the American Statistical Assn., 75, 122-132.

Woodfield, Terry J. (1982). Statistical Modeling of Bivariate Data. Ph.D. Thesis. Institute of Statistics, Texas A&M University.